

National Heart, Lung, and Blood Institute Workshop
The Promise of NHLBI Data Science
July 20 -21 2021

Workshop Agenda

Objectives

This workshop will be an informational and interactive opportunity for early, mid, and late-stage investigators, as well as graduate students, to address the following:

1. Knowledge gaps in understanding and utilizing NHLBI health datasets (e.g., BioData Catalyst)
2. Value of collaborations between domain experts and computer scientists, engineers and statisticians
3. Needs assessment to ensure diverse participation in HLBS data science

This will be accomplished through didactic sessions with dataset experts, interactive sessions and keynote speakers from health and computer science leaders. Collaborations encouraged during this meeting will be critical to bringing together the fields of health and computer science research.

DAY 1 (Zoom and archived)

SESSION 1 CHAIR: JONATHAN KALTMAN, M.D.

- | | |
|---------------------|---|
| 10:00 – 10:10 AM | Welcome and Opening Remarks
Gary Gibbons , M.D.
Director, National Heart, Lung, and Blood Institute |
| 10:10 – 10:20 AM | NHLBI Data Science Overview
David Goff , M.D., Ph.D.
Director, Division of Cardiovascular Sciences
National Heart, Lung, and Blood Institute |
| 10:20 – 10:45 AM | Keynote Address: Trustworthy AI Systems and Role of Datasets & Data Scientists
Danda Rawat, Ph.D.
Director, Data Science and Cybersecurity Center (DSC2)
Professor, Electrical Engineering and Computer Science
Howard University |
| 10:45 AM - 12:00 PM | Presentations on Datasets
Getting on NHLBI BioData Catalyst Powered by Seven Bridges
Dave Roberson, B.S.
Community Engagement Manager, Biomedical Research Platforms
Seven Bridges

Community Engagement for Biomedical Research Platforms
Seven Bridges
Alison Leaf , Ph.D.
Senior Program Manager
Seven Bridges |

12:00 - 1:00 PM Presentation: NHLBI-Generated Clinical and Genomic Big Data: Identify available genomic and clinical National Heart, Lung, Blood & Sleep Institute datasets and submit data access requests for analysis in the cloud.
[Sweta Ladwa](#), M.P.H., P.M.P.
Senior Scientific Program Manager, Information Technology and Application Center (ITAC)
National Heart, Lung, and Blood Institute

1:00 – 1:30 PM Lunch Break

SESSION 2 CHAIR: ERIN ITURRIAGA, D.N.P., M.S.N., R.N.

1:30 – 3:00 PM Introduction to Genome-Wide Association Studies (GWAS) resources in BioData Catalyst
[Beth Sheets](#), M.S.
Program Manager
UC Santa Cruz Genomics Institute

[Fayuan Wen](#), Ph.D.
Postdoctoral Associate
Howard University

3:00 – 4:00 PM Open Discussion Time

4:00 – 4:45 PM Getting Started on BioData Catalyst
[Amber Voght](#)
User Engagement Specialist
Renaissance Computing Institute at UNC (RENCI)

4:45 - 5:30 PM Wrap-up Day 1: Data Challenges Across Multiple Datasets and Novel Computational Methods
[Wendy Nilsen](#), Ph.D.
Program Director
Smart and Connected Health
Directorate for Computer & Information Science & Engineering
National Science Foundation

DAY 2 (Zoom and archived)

SESSION 3 CHAIR: ASIF RIZWAN, Ph.D.

10:00 - 10:30 AM Plenary Address: Towards Machine Learning for Personalized Healthcare
[Sanmi Koyejo](#), Ph.D.
Assistant Professor, Department of Computer Science
University of Illinois at Urbana-Champaign

10:30 – 11:15 AM Presentation: Application of Machine Learning and Artificial Intelligence Methods in Visualizing and Modelling of Complex Imaging and Clinical Data
[Xin Tian](#), Ph.D.
Mathematical Statistician, Division of Intramural Research
National, Heart, Lung, and Blood Institute

[Li-Yueh Hsu](#), D.Sc.

Staff Scientist, Radiology and Imaging Sciences
Clinical Center, National Institutes of Health

11:15 AM – 12:00 PM Presentation: Machine Learning Tools for Synergistically Mining Complex Data and Prior Knowledge

[George Em Karniadakis](#), Ph.D.

Professor of Applied Mathematics, Center for Fluid Mechanics
Brown University

12:00 – 12:45 PM Presentation: From Transcript to Tissue: Synthesis and Interpretation of Data Across Scales

[Jay Humphrey](#), Ph.D.

John C. Malone Professor of Biomedical Engineering
Department Chair, Biomedical Engineering
Yale University

Presentation: Interpreting Results from Genome-wide Searches – Experiences from TOPMed

[Ken Rice](#), Ph.D.

Professor, Department of Biostatistics
University of Washington

12:45 – 1:30 PM Lunch Break

SESSION 4 CHAIR: COLIN WU, Ph.D.

1:30 – 3:00 PM Case Study Presentations (BioData Catalyst Fellows)

Case Study 1 – Blood (genetic risk of allergic disease)

[Michelle Daya](#), Ph.D.

University of Colorado Denver

Project: HLA and Genome-Wide Association Studies of Total Serum IgE Levels

Case Study 2 – Heart (atrial fibrillation)

[Seung Hoan Choi](#), Ph.D.

Broad Institute of MIT and Harvard

Project: Genetic Architecture and Contribution of Rare Mutations to Atrial Fibrillation Risk

Case Study 3 – COPD (imaging phenotypes)

[Dandi Qiao](#), Ph.D.

Brigham and Women's Hospital

Project: Whole Genome-Sequencing Analyses of Imaging Phenotypes of Chronic Obstructive Pulmonary Disease (COPD)

Case Study 4 – Sickle Cell Disease (iron overload)

[Fayuan Wen](#), Ph.D.

Howard University

Project: Association Study of Iron Overload in Sickle Cell Disease Population Using NHLBI WGS from TOPMed

3:00 – 3:30 PM Plenary Address: National Science Board vision 2030 -The Importance of Diversity in STEM

[Victor McCrary](#), Jr., Ph.D.

Vice President, Research and Graduate Programs, University of the District of Columbia
Vice Chair, National Science Board

3:30 – 4:30PM Panel Discussion: Future Needs and Directions of Data Science in HBLs Research

[Jonathan Kaltman](#), M.D.

Senior Scientific Advisor/Lead in Data Science
National Heart, Lung, and Blood Institute

[Asif Rizwan](#), Ph.D.

Program Officer
Division of Blood Diseases and Resources
National Heart, Lung, and Blood Institute

[Colin Wu](#), Ph.D.

Program Officer/Math Statistician
Office of Biostatistics Research
National Heart, Lung, and Blood Institute

4:30 – 4:45 PM

Wrap Up for Organizers

[Erin Iturriaga](#), D.N.P., M.S.N., R.N.

Program Officer/Clinical Trials Specialist
Atherothrombosis and Coronary Artery Disease Branch
Division of Cardiovascular Sciences
National Heart, Lung, and Blood Institute